PCT/GB 2003 / 0 0 1 3 4 3

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

RECD 20 JUN 2003
WIPO PCT

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Stephen Hordley

Dated 9 April 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)

1/77

28NOV02 E766928-7 D00239
P01/7700 0.00-0227734.1

THE PATENT OFFICE
M
28 NOV 2002
NEWPORT

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference PC/JM/P12579GB

2. Patent application number *(The Patent Office will fill in this part)* 0227734.1

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

UNIVERSITY COURT OF THE UNIVERSITY OF EDINBURGH
Old College
South Bridge
Edinburgh EH8 9YL, UNITED KINGDOM

Patents ADP number *(if you know it)* 8858270001

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention Schizophrenia Associated Gene (VI)

5. Name of your agent *(if you have one)* CRUIKSHANK & FAIRWEATHER

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

19 Royal Exchange Square
Glasgow G1 3AE
UNITED KINGDOM

Patents ADP number *(if you know it)* 547002 ✓

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	24 ✓
Claim(s)	–
Abstract	–
Drawing(s)	6 + 6 ✓

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Cruikshank & Fairweather Date 29/11/02

CRUIKSHANK & FAIRWEATHER 27 NOVEMBER 2002

12. Name and daytime telephone number of person to contact in the United Kingdom

DR. P. CHAPMAN - 0141-221-5767

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

SCHIZOPHRENIA ASSOCIATED GENE (VI)

The present invention relates to the identification of a genes which has been disrupted in two patients diagnosed as suffering from schizophrenia and schizoaffective disorder, as well as proteins encoded by the genes and antibodies thereto and to uses of such products as medicaments for treating schizophrenia and/or affective psychosis. The invention also relates to methods for diagnosing patients suffering or predisposed to schizophrenia and/or affective psychosis, as well as screens for developing novel treatment regimes for schizophrenia and/or affective psychosis, such as schizophrenia.

Schizophrenia and Bipolar Affective Disorder are common and debilitating psychiatric disorders. Despite a wealth of information on the epidemiology, neuroanatomy and pharmacology of the illness, it is uncertain what molecular pathways are involved and how impairments in these affect brain development and neuronal function. Despite an estimated heritability of 60-80%, very little is known about the number or identity of genes involved in these psychoses. Although there has been recent progress in linkage and association studies, especially from genome-wide scans, these studies have yet to progress from the identification of susceptibility loci or candidate genes to the full characterisation of disease-causing genes (Berrettini, 2000).

The cloning of breakpoints in patients with chromosome abnormalities (translocations, inversions etc.) has proved instrumental in the identification of many disease genes (e.g. Duchenne Muscular Dystrophy, Retinoblastoma, Wilm's Tumour, Familial Polyposis Coli, Fragile-X Syndrome, Polycystic Kidney Disease, many leukaemias and, very recently, a candidate speech and language disorder gene (Lai et al, 2001)). Such studies assume that the chromosomal breakpoints give rise to the clinical symptoms by either directly disrupting

gene sequences or perturbing gene expression. In the same way that gene-trap mutagenesis can be used to identify disrupted mouse genes (Brennan & Skarnes, 1999), the physical "flag" created by a cytogenetic breakpoint provides a geographical pointer for the disease locus.

It is amongst the objects of the present invention to provide a gene and/or protein postulated to be involved with the development and/or symptoms associated with schizophrenia and/or affective psychosis.

As will be seen, the present invention is based on the molecular characterisation of a chromosomal rearrangement denoted t(1;16)(p31.2;q21).
The proband met ICD-10 and DSM-IV criteria for definite schizophrenia. The translocation was inherited within other branches of the family with variable clinical expression. However some key translocation carriers of the subjects to whom the inventors had access had not passed the age of risk when clinically characterized.

A high-throughput Fluorescence *in situ* Hybridisation (FISH)-based approach was adopted to map the chromosomal breakpoints in this patient. Consultation of the sequence data at the breakpoint loci not only allowed efficient FISH probe selections to be made by the targeting of coding regions, but also proof of gene disruption was inferred entirely by relating the exact position of probes to the genomic structure of a candidate gene.

One breakpoint (located on chromosome 1p31.2) in the subject lies within an alternatively spliced form of the gene, *PDE4B,* involved in the attenuation of cAMP secondary messenger signaling. This gene is the subject of filed patent GB0207900.2.

The remaining breakpoint in this patient (16q21) has now been fully characterised and demonstrated to disrupt a gene, *CADHERIN 8* (*CDH8*). The present invention is therefore based on a proposed role of this

gene in the aetiology of schizophrenia and/or affective psychosis.

Thus, in a first aspect the present invention provides use of a polynucleotide fragment comprising the *SEMCAP3* gene, or a fragment, derivative or homologue thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

In another aspect, the present invention provides use of a polypeptide fragment encoded by the *CDH8* gene, or fragments, derivatives or homologues thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

Schizophrenia and/or affective psychosis as used herein relates to schizophrenia, as well as other affective psychoses such as those listed in "The ICD-10 Classification of Mental and Behavioural Disorders" World Health Organization, Geneva 1992. Categories F20 to F29 inclusive includes Schizophrenia, schizotypal and delusional disorders. Categories F30 to F39 inclusive are Mood (affective) disorders that include bipolar affective disorder and depressive disorder. Mental Retardation is coded F70 to F79 inclusive. The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). American Psychiatric Association, Washington DC. 1994. Include all conditions coded 295.xx (Schizophrenia and Other Psychotic Disorders) and 296.xx (Major Depressive Disorders and Bipolar Disorders). Mental Retardation is coded 315, 317, 318 and 319.

CADHERIN 8 (*CDH8*) has been previously cloned and sequenced and the sequence is present in the public database (nucleic acid sequence; L34060/AB035305/NM_001796, protein sequence; NP_001787) and described in Suzuki *et al.*, 1991, Tanihara *et al.*, 1994, and Shimoyama *et al.*, 2000. An alternative transcript form has been described in the rat in which there is a truncation within the 5th cadherin domain (Kido et al., 1998 and see Fig.4). The accession

numbers for the normal and truncated forms of *CDH8* in rat are AB010436 and AB010437, respectively. The corresponding human truncated transcript is not present in the public database and so is not yet confirmed. The genomic sequences corresponding to *CDH8* are also present in the public database (eg. BAC CTC-420A11; AC040161). Nevertheless, the prior art does not suggest any link between *CDH8* and schizophrenia and/or affective psychosis.

Thus, references herein to the *CDH8* gene are understood to relate to the nucleic sequences in the public databases and identified in Fig.3 and references to the CDH8 protein sequences are understood to relate to the sequences in the public databases and identified in Fig.4.

In certain jurisdictions claims to methods of treatment are permissible and so the skilled reader will appreciate that the *CDH8* gene, or fragments, derivatives or homologues thereof; or CDH8 proteins, or functionally active fragments, derivatives, or homologues thereof, may be administered to an individual as a method of treating an individual with schizophrenia and/or affective psychosis.

"Polynucleotide fragment" as used herein refers to a chain of nucleotides such as deoxyribose nucleic acid (DNA) and transcription products thereof, such as RNA. Naturally, the skilled addressee will appreciate the whole naturally occurring human genome is not included in the definition of polynucleotide fragment.

The polynucleotide fragment can be isolated in the sense that it is substantially free of biological material with which the whole genome is normally associated *in vivo*. The isolated polynucleotide fragment may be cloned to provide a recombinant molecule comprising the polynucleotide fragment. Thus, "polynucleotide fragment includes double and single stranded DNA, RNA and polynucleotide sequences derived therefrom, for example, subsequences of said fragment and which are of any desirable length. Where a nucleic

acid is single stranded then both a given strand and a sequence or reverse complementary thereto is within the scope of the present invention.

In general, the term "expression product" or "gene product" refers to both transcription and translation products of said polynucleotide fragments. When the expression or gene product is a "polypeptide" (i.e. a chain or sequence of amino acids displaying a biological activity substantially similar (eg. 98%, 95%, 90%, 80%, 75% activity) to the biological activity of the protein), it does not refer to a specific length of the product as such. Thus, the skilled addressee will appreciate that "polypeptide" encompasses *inter alia* peptides, polypeptides and proteins. The polypeptide if required, can be modified *in vivo* and *in vitro*, for example by glycosylation, amidation, carboxylation, phosphorylation and/or post-translational cleavage.

The present invention further provides a recombinant or synthetic polypeptide for the manufacture of reagents for use as therapeutic agents in the treatment of schizophrenia and/or affective psychosis. In particular, the invention provides pharmaceutical compositions comprising the recombinant or synthetic polypeptide together with a pharmaceutically acceptable carrier therefor.

The present invention further provides an isolated polynucleotide fragment capable of specifically hybridising to a related polynucleotide sequence from another species. In this manner, the present invention provides probes and/or primers for use in *ex vivo* and/or *in situ* detection and expression studies. Typical detection studies include polymerase chain reaction (PCR) studies, hybridisation studies, or sequencing studies. In principle any specific polynucleotide sequence fragment from the identified sequences may be used in detection and/or expression studies. The skilled addressee understands that a specific fragment is a fragment of the sequence which

is of sufficient length, generally greater than 10, 12, 14, 16 or 20 nucleotides in length, to bind specifically to the sequence, under conditions of high stringency, as defined herein, and not bind to unrelated sequences, that is sequences from elsewhere in the genome of the organism other than an allelic form of the sequence or non-homologous sequences from other organisms.

"Capable of specifically hybridising" is taken to mean that said polynucleotide fragment preferably hybridises to a related or similar polynucleotide sequence in preference to unrelated or dissimilar polynucleotide sequences.

The invention includes polynucleotide sequence(s) which are capable of specifically hybridising to a polynucleotide fragment as described herein or to a part thereof without necessarily being completely complementary or reverse complementary to said related polynucleotide sequence or fragment thereof. For example, there may be at least 50%, or at least 75%, at least 90%, or at least 95% complementarity. Of course, in some cases the sequences may be exactly reverse complementary (100% reverse complementary) or nearly so (e.g. there may be less than 10, typically less than 5 mismatches). Thus, the present invention also provides anti-sense or complementary nucleotide sequence(s) which is/are capable of specifically hybridising to the disclosed polynucleotide sequence. If a specific polynucleotide is to be used as a primer in PCR and/or sequencing studies, the polynucleotide must be capable of hybridising to related nucleic acid and capable of initiating chain extension from 3' end of the polynucleotide, but not able to correctly initiate chain extension from unrelated sequences.

If a polynucleotide sequence of the present invention is to be used in hybridisation studies to obtain or identify a related sequence from another organism the polynucleotide sequence should preferably remain hybridised to a sample polynucleotide under

stringent conditions. If desired, either the test or sample polynucleotide may be immobilised. Generally the test polynucleotide sequence is at least 10, 14, 20 or at least 50 bases in length. It may be labelled by suitable techniques known in the art. Preferably the test polynucleotide sequence is at least 200 bases in length and may even be several kilobases in length. Thus, either a denatured sample or test sequence can be first bound to a support. Hybridization can, for example, be effected at a temperature of between 50 and 70°C in double strength SSC (2xNaCl 17.5g/l and sodium citrate (SC) at 8.8g/l) buffered saline containing 0.1% sodium dodecyl sulphate (SDS). This can be followed by rinsing of the support at the same temperature but with a buffer having a reduced SSC concentration. Depending upon the degree of stringency required, and thus the degree of similarity of the sequences, such reduced concentration buffers are typically single strength SSC containing 0.1%SDS, half strength SSC containing 0.1%SDS and one tenth strength SSC containing 0.1%SDS. Sequences having the greatest degree of similarity are those the hybridisation of which is least affected by washing in buffers of reduced concentration. It is most preferred that the sample and inventive sequences are so similar that the hybridisation between them is substantially unaffected by washing or incubation in standard sodium citrate (0.1 x SSC) buffer containing 0.1%SDS.

Oligonucleotides may be designed to specifically hybridise to *CDH8* nucleic acids. They may be synthesised, by known techniques and used as primers in PCR or sequencing reactions or as probes in hybridisations designed to detect the presence of a mutated or normal *CDH8* genes in a sample. The oligonucleotides may be labelled by suitable labels known in the art, such as, radioactive labels, chemiluminescent labels or fluorescent labels and the like. The term "oligonucleotide" is not meant to indicate any particular length of sequence and

encompasses nucleotides of preferably at least 10b (e.g. 10b to 1kb) in length, more preferably 12b-500b in length and most preferably 15b to 100b.

The oligonucleotides may be designed with respect to any of the sequences described herein and may be manufactured according to known techniques. They may have substantial sequence identity (e.g. at least 50%, at least 75%, at least 90% or at least 95% sequence identity) with one of the strands shown herein or an RNA equivalent, or with a part of such a strand. Preferably such a part is at least 10, at least 30, at least 50 or at least 200 bases long. It may be an open reading frame (ORF) or a part thereof.

Oligonucleotides which are generally greater than 30 bases in length should preferably remain hybridised to a sample polynucleotide under one or more of the stringent conditions mentioned above. Oligonucleotides which are generally less than 30 bases in length should also preferably remain hybridised to a sample polynucleotide but under different conditions of high stringency. Typically the melting temperature of an oligonucleotide less than 30 bases may be calculated according to the formula of; 2°C for every A or T, plus 4°C for every G or C, minus 5°C. Hybridization may take place at or around the calculated melting temperature for any particular oligonucleotide, in 6 x SSC and 1% SDS. Non specifically hybridised oligonucleotides may then be removed by stringent washing, for example in 3 x SSC and 0.1% SDS at the same temperature. Only substantially similar matched sequences remain hybridised i.e. said oligonucleotide and corresponding test nucleic acid.

When oligonucleotides of generally less than 30 bases in length are used in sequencing and/or PCR studies, the melting temperature may be calculated in the same manner as described above. The oligonucleotide may then be allowed to anneal or hybridise at a temperature around the oligonucleotides calculated melting temperature. In the case of PCR

studies the annealing temperature should be around the lower of the calculated melting temperatures for the two priming oligonucleotides. It is to be appreciated that the conditions and melting temperature calculations are provided by way of example only and are not intended to be limiting. It is possible through the experience of the experimenter to vary the conditions of hybridisation and thus anneal/hybridise oligonucleotides at temperatures above their calculated melting temperature. Indeed this can be desirable in preventing so-called non-specific hybridisation from occurring.

It is possible when conducting PCR studies to predict an expected size or sizes of PCR product(s) obtainable using an appropriate combination of two or more oligonucleotides, based on where they would hybridise to the sequences described herein. If, on conducting such a PCR on a sample of DNA, a fragment of the predicted size is obtained, then this is predictive that the DNA encodes a homologous sequence from a test organism.

Proteins for all the applications described herein can be produced by cloning the gene for example into plasmid vectors that allow high expression in a system of choice e.g. insect cell culture, yeast, animal cells, bacteria such as *Escherichia coli*. To enable effective purification of the protein, a vector may be used that incorporates an epitope tag (or other "sticky" extension such as His6) onto the protein on synthesis. A number of such vectors and purification systems are commercially available.

The polynucleotide fragment can be molecularly cloned into a prokaryotic or eukaryotic expression vector using standard techniques and administered to a host. The expression vector is taken up by cells and the polynucleotide fragment of interest expressed, producing protein.

It will be understood that for the particular polypeptides embraced herein, natural variations such

as may occur due to polymorphisms, can exist between individuals or between members of the family. These variations may be demonstrated by (an) amino acid difference(s) in the overall sequence or by deletions, substitutions, insertions, inversions or additions of (an) amino acid(s) in said sequence. All such derivatives showing the recognised activity are included within the scope of the invention. For example, for the purpose of the present invention conservative replacements may be made between amino acids within the following groups:

(I) Alanine, serine, threonine;

(II) Glutamic acid and aspartic acid;

(III) Arginine and leucine;

(IV) Asparagine and glutamine;

(V) Isoleucine, leucine and valine;

(VI) Phenylalanine, tyrosine and tryptophan

Moreover, recombinant DNA technology may be used to prepare nucleic acid sequences encoding the various derivatives outlined above.

As is well known in the art, the degeneracy of the genetic code permits substitution of bases in a codon resulting in a different codon which is still capable of coding for the same amino acid, e.g. the codon for amino acid glutamic acid is both GAT and GAA. Consequently, it is clear that for the expression of polypeptides from nucleotide sequences described herein or fragments thereof, use can be made of a derivative nucleic acid sequence with such an alternative codon composition different from the nucleic acid sequences shown in the Figures.

The polynucleotide fragments of the present invention are preferably linked to regulatory control sequences. Such control sequences may comprise promoters, operators, inducers, enhancers, silencers, ribosome binding sites, terminators etc. Suitable control sequences for a given host may be selected by those of ordinary skill in the art.

A polynucleotide fragment according to the present invention can be ligated to various expression controlling sequences, resulting in a so called recombinant nucleic acid molecule. Thus, the present invention also includes an expression vector containing an expressible nucleic acid molecule. The recombinant nucleic acid molecule can then be used for the transformation of a suitable host.

Specific vectors which can be used to clone nucleic acid sequences according to the invention are known in the art (e.g. Rodriguez, R.L. and Denhadt, D.T., Edit., Vectors: a survey of molecular cloning vectors and their uses, Butterworths, 1988, or Jones *et al.*, Vectors: Cloning Applications: Essential Techniques (Essential techniques series), John Wiley & Son. 1998).

The methods to be used for the construction of a recombinant nucleic acid molecule according to the invention are known to those of ordinary skill in the art and are inter alia set forth in Sambrook, et al. (Molecular Cloning: a laboratory manual Cold Spring Harbour Laboratory, 1989).

The present invention also relates to a transformed cell containing the polynucleotide fragment in an expressible form. "Transformation", as used herein, refers to the introduction of a heterologous polynucleotide fragment into a host cell. The method used may be any known in the art, for example, direct uptake, transfection transduction or electroporation (Current Protocols in Molecular Biology, 1995. John Wiley and Sons Inc.). The heterologous polynucleotide fragment may be maintained through autonomous replication or alternatively, may be integrated into the host genome. The recombinant nucleic acid molecules preferably are provided with appropriate control sequences compatible with the designated host which can regulate the expression of the inserted polynucleotide fragment, e.g. tetracycline responsive

promoter, thymidine kinase promoter, SV-40 promoter and the like.

Suitable hosts for the expression of recombinant nucleic acid molecules may be prokaryotic or eukaryotic in origin. Hosts suitable for the expression of recombinant nucleic acid molecules may be selected from bacteria, yeast, insect cells and mammalian cells.

In another aspect, the present invention also relates to a method of diagnosing schizophrenia and/or affective psychosis or susceptibility to schizophrenia and/or affective psychosis in an individual, wherein the method comprises determining if the *CDH8* gene in the individual has been disrupted by a mutation or chromosomal rearrangement.

The methods which may be employed to elucidate such a mutation or chromosomal rearrangement are well known to those of skill in the art and could be detected for example using PCR or in hybridisation studies using suitable probes which could be designed to span an identified mutation site or chromosomal breakpoint in close proximity to the *CDH8* gene, such as the breakpoint identified by the present inventors and described herein.

Once a particular polymorphism or mutation has been identified it may be possible to determine a particular course of treatment. For example it is known that some forms of treatment work for some patients, but not all. This may in fact be due to mutations in the *CDH8* gene or surrounding sequences, and it may therefore be possible to determine a treatment strategy using current therapies, based on a patient's genotype.

It will be appreciated that mutations in the gene sequence or controlling elements of a gene, eg. a promoter and/or enhancer can have subtle effects such as affecting mRNA splicing/stability/activity and/or control of gene expression levels, which can also be determined. Also the relative levels of RNA can be determined using for example hybridisation or

quantitative PCR as a means to determine if the *CDH8* gene has been disrupted.

Moreover the presence and/or levels of the *CDH8* gene products themselves can be assayed by immunological techniques such as radioimmunoassay, Western blotting and ELISA using specific antibodies raised against the gene products. The present invention also therefore relates to antibodies specific for *CDH8* gene products and uses thereof in diagnosis and/or therapy.

Thus in a further aspect of the present invention therefore provides antibodies specific to the polypeptides of the present invention or epitopes thereof. Production and purification of antibodies specific to an antigen is a matter of ordinary skill, and the methods to be used are clear to those skilled in the art. The term antibodies can include, but is not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanised or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope binding fragments of any of the above. Such antibodies may be used in modulating the expression or activity of the particular polypeptide, and/or in detecting said polypeptide *in vivo* or *in vitro* and thus used, for example, in the treatment and/or diagnosis of schizophrenia and/or affective psychosis.

Using the sequences disclosed herein, it is possible to identify related sequences in other animals, such as mammals, with the intention of providing an animal model for psychiatric disorders associated with the improper functioning of the nucleotide sequences and proteins of the present invention. Once identified, the homologous sequences can be manipulated in several ways common to the skilled person in order to alter the functionality of the nucleotide sequences and proteins homologous to those of the present invention. For example, "knock-

out" animals may be created, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be reduced or substantially eliminated in order to determine the effects of reducing or substantially eliminating the expression of such genes. Alternatively, animals may be created where the expression of the nucleotide sequences and proteins homologous to those of the present invention are upregulated, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be increased in order to determine the effects of increasing the expression of these genes. In addition to these manipulations, substitutions, deletions and additions may be made to the nucleotide sequences encoding the proteins homologous to those of the present invention in order to effect changes in the activity of the proteins to help elucidate the function of domains, amino acids, etc. in the proteins. Furthermore, the sequences of the present invention may also be used to transform animals to the manner described above. The manipulations described above may also be used to create an animal model of schizophrenia and/or affective psychosis associated with the improper functioning of the nucleotide sequences and/or proteins of the present invention in order to evaluate potential agents which may be effective for combatting psychotic disorders, such as schizophrenia and/or affective psychosis.

Thus, the present invention also provides for screens for identifying agents suitable for preventing and/or treating schizophrenia and/or affective psychosis associated with disruption or alteration in the expression of the *CDH8* gene and/or their gene products. Such screens may easily be adapted to be used for the high throughput screening of libraries of compounds such as synthetic, natural or combinatorial compound libraries.

Thus, the *CDH8* gene products according to the present invention can be used for the *in vivo* or *in vitro* identification of novel ligands or analogs thereof. For this purpose binding studies can be performed with cells transformed with nucleotide fragments according to the invention or an expression vector comprising a polynucleotide fragment according to the invention, said cells expressing the *CDH8* gene products according to the invention.

Alternatively also the *CDH8* gene products according to the invention as well as ligand-binding domains thereof can be used in an assay for the identification of functional ligands or analogs for the *CDH8* gene products.

Methods to determine binding to expressed gene products as well as *in vitro* and *in vivo* assays to determine biological activity of gene products are well known. In general, expressed gene product is contacted with the compound to be tested and binding, stimulation or inhibition of a functional response is measured.

Thus, the present invention provides for a method for identifying ligands for *CDH8* gene products, said method comprising the steps of:

a) introducing into a suitable host cell a polynucleotide fragment according to the invention;

b) culturing cells under conditions to allow expression of the polynucleotide fragment;

c) optionally isolating the expression product;

d) bringing the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said polynucleotide fragment from step a);

e) establishing whether a ligand has bound to the expressed protein; and

f) optionally isolating and identifying the ligand.

As a preferred way of detecting the binding of the ligand to the expressed protein, also signal transduction capacity may be measured.

Compounds which activate or inhibit the function of *CDH8* gene products may be employed in therapeutic treatments to activate or inhibit the polypeptides of the present invention.

The present invention will now be further described by way of Example and with reference to the Figures which show:

Figure 1 shows an ideogram diagram of the chromosomal rearrangement (a reciprocal translocation) in this patient. The two breakpoints are marked at the approximate chromosomal locations at which they are located. In addition, and not to scale, the two candidate disease-causing genes, *PDE4B* and *CDH8*, are placed in the correct orientation and with respect to the breakpoints. The fusion genes on derived chromosomes 1 and 16 that result from the reciprocal translocation are also indicated, demonstrating the potential capacity for fusion transcript/protein synthesis.

Figure 2 shows a representation of the genomic structure of the *CDH8* gene: its spliced exons spread over a genomic extent of approximately 400kb.
Above the gene, the coding contribution of each exon to the CDH8 protein is indicated by bars and finely dashed lines. The domain structure of CDH8 protein is shown at the top of the figure. `N' and `C' refer to the N- and C-termini of the protein. The broken line at the N-terminus indicates the existence of signal peptide and proprotein domains - both of which are cleaved off in the mature protein. The `CD' ovals represent the positions of the five extracellular cadherin domains. The black box signifies the position of the hydrophobic stretch of amino acids that act as the membrane-spanning domain. The BAC clones used to identify the breakpoint location are included at the bottom of the figure together with the inferred direction (arrows) of the breakpoint from the FISH results using these clones. The heavy dashed line shows the position of the

breakpoint with respect to the gene exons and the domain structure of the protein.

Figure 3 Nucleic acid sequence of Human *CDH8*. The following features are marked for clarity:

a) ATG start site located at position 253 (underlined)

b) GC bases (underlined) at the junction between exons 1 and 2 (i.e. between which the breakpoint is located)

c) UGA stop codon located at position 2650 (underlined).

Figure 4 Amino acid sequence of Human CDH8 with underlined regions of interest.

a) Residues 1-29 signal peptide domain (italics)

b) Residues 30-61 propeptide fragment cleaved off in mature protein.

c) Residues 76-158 cadherin domain #1 (underlined)

d) Residues 172-248 cadherin domain #2 (underlined)

e) Residues 281-383 cadherin domain #3 (underlined)

f) Residues 396-487 cadherin domain #4 (underlined)

g) Residues 500-597 cadherin domain #5 (underlined)

h) `V' highlighted at position 513 is the last residue in common with the putative truncated rat protein product from the alternatively spliced form.

i) Residues 622-645 transmembrane domain #1 (underlined).

Figure 5

a) Fusion protein product resulting from CDH8 promoter/exon 1 spliced to *PDE4B* exon 2 and beyond (transcribed on der(16)). The underlined residues `RV' represent the fusion site between the two genes.

b) Fusion protein product resulting from *PDE4B* promoter (long form)/exon 1a spliced to *CDH8* exon 2 and beyond (transcribed on der(1)). See text for details: only the reading frame producing the N-terminal truncated form of the CDH8 protein is shown. The underlined `gc' at position 68 represents the point of fusion between the two genes. Three potential methionine translation start sites are

shown (highlighted) with the second of these having a nucleic acid sequence most similar to the canonical Kozak sequence (underlined). Use of this start site would generate a truncated CDH8 protein lacking the signal peptide, proprotein fragment, cadherin domain 1 and most of cadherin domain 2.

Materials and methods

Lymphocyte extraction and metaphase chromosome preparation

Lymphocytes were extracted from 7mls of patient blood (for storage and generation of EBV-transformed cell lines) using density gradient separation (Histopaque-1077, Sigma). In order to generate metaphase-arrested chromosomes for cytogenetic analysis, 0.8mls of patient blood were cultured for 71hrs in medium containing phytohaemagglutinin (Peripheral Blood Medium, Sigma). The short-term cultures were treated with colcemid for one hour followed by a conventional fixing procedure. Fixed chromosomes were dropped onto microscope slides and stored for 1 week prior to use in FISH experiments.

Fluorescence in situ hybridisation (FISH) protocol

Probe template DNA (e.g BAC clone DNA) was labelled by nick translation and hybridised to patient metaphase spreads using standard FISH methods. Slides were counterstained with DAPI in Vectashield anti-fade solution (Vector laboratories). A Zeiss Axioskop fluorescence microscope with a chroma number 83000 multi-spectral filter set was used to observe the chromosomal hybridisations. Images were captured using Digital Scientific SmartCapture imaging software. FISH signals observed on derived chromosomes dictated the selection of further clones required to "walk" towards the breakpoint. Breakpoint-spanning FISH probes have signals on a normal chromosome and on both derived chromosomes.

Resolution of breakpoint position

BAC clones corresponding to proposed breakpoint regions were arranged into contigs by consulting the Washington University FPC (http://www.genome.wustl.edu/gsc/human/Mapping/index.shtml), UCSC GoldenPath Draft Human Genome Browser (http://genome.cse.ucsc.edu/index.html?org=Human) and Ensembl (http://www.ensembl.org/) databases. BAC clones were supplied by BACPAC Resources, Oakland, California, USA (http://www.chori.org/bacpac/). Clone selection was biased to gene-containing BACs.

Example 1: Molecular characterisation of chromosomal disruption and identification of disrupted gene

FISH experiments on chromosome 16q21 had narrowed the location of the breakpoint to a region including the large gene *CDH8* (approximately 400kb genomic extent). Three BAC clones were selected from the tiling diagram of BAC clones placed on the human genome map backbone (June 2002 release of the 'BAC End Pairs' track on the UCSC Genome Browser; http://genome.cse.ucsc.edu/index.html?org=Human). These were RPCI-11 599c11, RPCI-11 875e12 and RPCI-11 685m21. By FISH, these BAC clones flanked the breakpoint (the first two translocated the derived chromosome 1 whereas the third remained on the derived chromosome 16). The position of these three BAC clones indicated that the breakpoint lay within the large (100kb) intron between exons 1 and 2 of the *CDH8* gene (see Fig.2). Thus, the inventors inferred from these results that the *CDH8* gene was directly disrupted by the 16q21 translocation event and, as such, is a candidate gene for the psychiatric disorder exhibited by the patient. The similar disruption of the *PDE4B* gene on chromosome 1 (GB0207900.2) and their relative orientations on the two chromosomes raised the possibility that the derived chromosomes (the two chromosomes resulting from the translocation: der(1) and der(16)) could transcribe fusion/hybrid genes. This has been frequently seen in

cases where a translocation gives rise to susceptibilty to cancers. In essence, the translocation in the proband resulted in an exchange of the two genes' promoter and first exon sequences. On the der(1) the promoter and first exon of the *CDH8* gene are juxtaposed to exon 2 and downstream of the *PDE4B* gene (see Fig.1). However, the reading frames of these two gene segments are not the same, resulting in a prematurely truncated peptide with only the signal peptide, proprotein fragment and a small portion of the cadherin domain contained within (see Fig. 5a). This would be expressed in the same cell types/tissues as the normal *CDH8* gene but the functional/pathological significance of this small peptide is not clear at the current time. On the der(16) the *PDE4B* promoter and exon 1a are juxtaposed to exon 2 and downstream of the *CDH8* gene (see Fig.1). Exon 1a of *PDE4B* does not contain a translation start-site so the reading frame compatabilities of the putative fused transcript are not an issue. However, exon 2 and downstream of the *CDH8* gene contain several ATG start-sites which could be employed by translational machinery to generate peptide sequences. In two of the reading frames, any generated peptides would be small and probably of no consequence. The third reading frame (the normal *CDH8* reading frame, see Fig.5b) contains three ATG start-sites early on, with the second of these forming a very good match to the canonical Kozak sequence found at most translation start-sites (CCAxx**ATG**G). If this one is used then the resulting peptide will be identical to normal CDH8 protein but lacking the N-terminal portion encoding the signal peptide, proprotein fragment, the first cadherin domain and most of the second cadherin domain. Although the bulk of the peptide sequence is as the normal CDH8 protein, the lack of the N-terminal sequences may prevent the protein from entering the Golgi/ER subcellular compartments - a process that is required for the correct insertion in/trafficking to the cell membrane. The functional/pathological consequence of

the presence of this truncated form of the CDH8 protein in the cytoplasm of tissues where the long form of the *PDE4B* gene is expressed is uncertain at this point.
In summary, the psychiatric illness seen in the proband, and other members of the family, may be the result of one (or a combination) of the following circumstances: the loss (through disruption) of one allele of *PDE4B*, the loss (through disruption) of one allele of *CDH8* or the generation of potentially pathological fusion polypeptides.

Cadherin-8 was first cloned in humans (Tanihara et al., 1994) and later in mouse (Munro et a., 1996) and rat (Kido et al., 1998). Sequence analysis immediately placed the gene product within the large family of membrane-spanning proteins with extracellular cadherin domains thought to mediate calcium-dependent homophilic interactions between adjacent cells. As such, the cadherins are members of the functionally defined group of cell adhesion proteins.

CDH8 is a member of the Type II, or atypical, cadherins which are defined by the lack of an extracellular tripeptide motif, HAV, possibly involved in the binding specificity of Type I cadherins. Fig.2 illustrates the structure of CDH8 protein which includes an extracellular domain containing 5 copies of the cadherin domain, a membrane spanning domain and a C-terminal cytoplasmic tail. The cytoplasmic tail is thought to signal the presence of interactions to the intracellular compartment by mediating receptor clustering through interaction with the proteins such as β-catenin, α-catenin and, eventually, the cytoskeletal proteins, actin and α-actinin. In this way, adhesion to adjacent cells can affect the cytoarchitecture of the cell and may even play a role in cell motility.

The two principal roles of neuronal cadherins are thought to be in the mediation of certain developmental pathways in the brain and the regulation of synaptic

function. The homophilic nature of cadherin interaction (i.e. CDH8 proteins preferentially bind to other CDH8 proteins) has prompted the hypothesis that cadherins are responsible for the aggregation or interconnection of similar cells within an organ. This has been shown to be the case in the brain where *CDH8* expression has been shown to be restricted to particular subregions and even neuronal patches (Redies, Bishop, Rubenstein, Korematsu X 2).

The major cadherin in the brain, N-cadherin (encoded by *CDH2*), has been implicated in synaptic long-term potentiation (LTP): the mechanism thought to underlie learning and memory on the brain (e.g. Huntley et al., 2002 & Bozdagi et al., 2000,). Other cadherins may also play a part in this process (Uemura, 1998 & Tang et al., 1998). In essence, cadherins seem to form physical bridges across the synaptic cleft which may modify synaptic efficacy and/or spine morphology (two features of neurons demonstrated to change after the induction of LTP).

Interestingly, two of the hypotheses used to explain the origins of psychiatric illness are, firstly, the occurrence of abnormal brain development and, secondly, the existence of deficits in cellular pathways manifested as poor performance in certain cognitive/memory tasks. The two roles of neuronal cadherins seem to closely mirror these two hypotheses suggesting that CDH8 is a good functional candidate for psychiatric illness.

REFERENCES

Berrettini W.H. (2000) Genetics of psychiatric disease. Annu. Rev. Med. 51:465 - 79.

Lai C.S, Fisher S.E, Hurst J.A, Vargha-Khadem F, Monaco A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 413(6855):519 - 23.

Kleinjan D.J, van Heyningen V. (1998) Position effect in human genetic disease. Hum Mol Genet. 7(10):1611 - 8.

Brennan J, Skarnes W.C. (1999) Gene trapping in mouse embryonic stem cells. Methods Mol. Biol. 97:123 - 38.

Breen M, Arveiler B, Murray I, Gosden J.R, Porteous D.J. (1992) YAC mapping by FISH using Alu-PCR-generated probes. Genomics. 13(3):726 - 30.

Wang LH, Kalb RG, Strittmatter SM. (1999) A PDZ protein regulates the distribution of the transmembrane semaphorin, M-SemF. J Biol Chem. 1999 274(20):14137-46.

Shimoyama Y, Tsujimoto G, Kitajima M, Natori M. (2000) Identification of three human type-II classic cadherins and frequent heterophilic interactions between different subclasses of type-II classic cadherins. Biochem J. 349(Pt 1):159-67.

Korematsu K, Goto S, Okamura A, Ushio Y. (1998) Heterogeneity of cadherin-8 expression in the neonatal rat striatum: comparison with striatal compartments. Exp Neurol.154(2):531-6.

Korematsu K, Nishi T, Okamura A, Goto S, Morioka M, Hamada J, Ushio Y. (1998) Cadherin-8 protein expression in gray matter structures and nerve fibers of the neonatal and adult mouse brain. Neuroscience. 87(1):303-15.

Kido M, Obata S, Tanihara H, Rochelle JM, Seldin MF, Taketani S, Suzuki ST. (1998) Molecular properties and chromosomal location of cadherin-8. Genomics. 1998 48(2):186-94.

Korematsu K, Redies C. (1997) Expression of cadherin-8 mRNA in the developing mouse central nervous system. J Comp Neurol.387(2):291-306.

Korematsu K, Redies C. (1997) Restricted expression of cadherin-8 in segmental and functional subdivisions of the embryonic mouse brain. Dev Dyn. 208(2):178-89.

Tanihara H, Sano K, Heimark RL, St John T, Suzuki S. (1994) Cloning of five human cadherins clarifies characteristic features of cadherin extracellular domain and provides further evidence for two structurally different types of cadherin. Cell Adhes Commun. 2(1):15-26.

Suzuki S, Sano K, Tanihara H. (1991) Diversity of the cadherin family: evidence for eight new cadherins in nervous tissue. Cell Regul. 2(4):261-70.

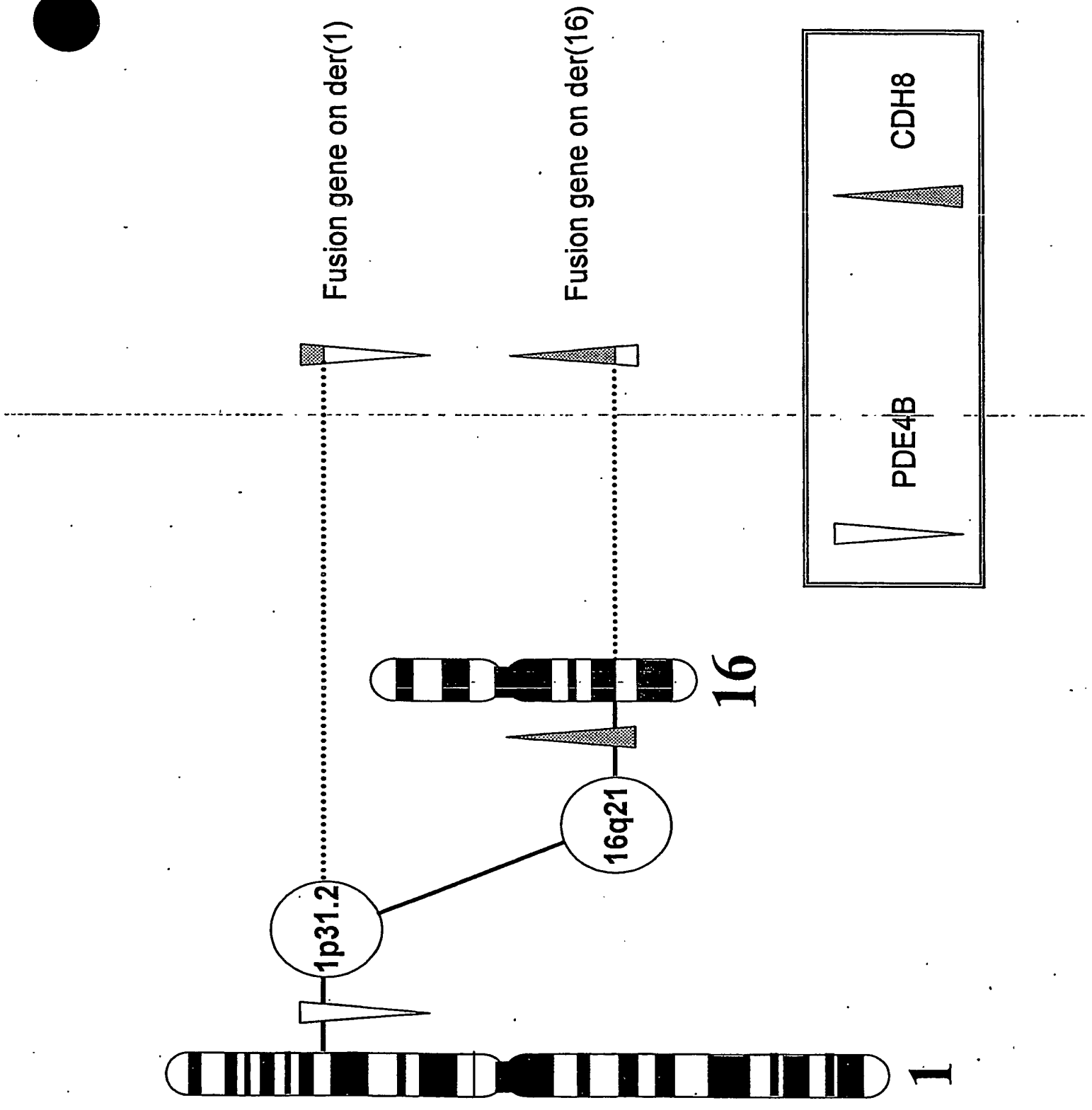
Fusion gene on der(1)
Fusion gene on der(16)
CDH8
PDE4B
16
16q21
1p31.2
1

Figure 1.

Figure 3

```
   1 agccatttgt gaacctggag gcttgacatt cgccagcgca gggccccaca agagaaattt
  61 caatgaaaag aaaagccaat ggattgtggt cttagaaaag ctgcttagat gatgtctgtt
 121 tcccgtgcta tagacacgtg gcagagctgt aagtaaatgc tcggcactgc atgatgaatt
 181 ggatggctgc agaccggaga caaaaaaaat aattgtctca ttttcgtggt gatttgctta
 241 actggtggga ccatgccaga acggctagcg gaaatgctct tggatctctg gactccatta
 301 ataatattat ggattactct tccccccttgc atttacatgg ctccgatgaa tcagtctcaa
 361 gttttaatga gtggatcccc tttggaacta aacagtctgg gtgaagaaca gcgaattttg
 421 aaccgctcca aaagaggctg ggtttggaat caaatgtttg tcctggaaga gttttctgga
 481 cctgaaccga ttcttgttgg ccggctacac acagacctgg atcctgggag caaaaaaatc
 541 aagtatatcc tatcaggtga tggagctggg accatatttc aaataaatga tgtaactgga
 601 gatatccatg ctataaaaag acttgaccgg gaggaaaagg ctgagtatac cctaacagct
 661 caagcagtgg actgggagac aagcaaacct ctggagcctc cttctgaatt tattattaaa
 721 gttcaagaca tcaatgacaa tgcaccagag tttcttaatg gaccctatca tgctactgtg
 781 ccagaaatgt ccattttggg tacatctgtc actaacgtca ctgcgaccga cgctgatgac
 841 ccagtttatg gaaacagtgc aaagttggtt tatagtatat tggaagggca gccttatttt
 901 tccattgagc ctgaaacagc tattataaaa actgcccttc ccaacatgga cagagaagcc
 961 aaggaggagt acctggttgt tatccaagcc aaagatatgg gtggacactc tggtggcctg
1021 tctgggacca cgacacttac agtgactctt actgatgtta atgacaatcc tccaaaattt
1081 gcacagagcc tgtatcactt ctcagtaccg gaagatgtgg ttcttggcac tgcaatagga
1141 agggtgaagg ccaatgatca ggatattggt gaaaatgcac agtcatcata tgatatcatc
1201 gatggagatg gaacagcact ttttgaaatc acttctgatg cccaggccca ggatggcatt
1261 ataaggctaa gaaaacctct ggactttgag accaaaaaat cctatacgct aaaggtagag
1321 gcagccaatg tccatattga cccacgcttc agtggcaggg ggccctttaa agacacggcg
1381 acagtcaaaa tcgtggttga agatgctgat gagcctccgg tcttctcttc accgacttac
1441 ctacttgaag ttcatgaaaa tgctgctcta aactccgtga ttgggcaagt gactgctcgt
1501 gaccctgata tcacttccag tcctataagg ttttccatcg accggcacac tgacctggag
1561 aggcagttca acattaatgc agacgatggg aagataacgc tggcaacacc acttgacaga
1621 gaattaagtg tatggcacaa cataacaatc attgctactg aaattaggaa ccacagtcag
1681 atatcacgag tacctgttgc tattaaagtg ctggatgtca atgacaacgc ccctgaattc
1741 gcatccgaat atgaggcatt tttatgtgaa aatggaaaac ccggccaagt cattcaaact
1801 gttagcgcca tggacaaaga tgatcccaaa aacggacatt atttcttata cagtctcctt
1861 ccagaaatgg tcaacaatcc gaatttcacc atcaagaaaa atgaagataa ttccctcagt
1921 attttggcaa agcataatgg attcaaccgc cagaagcaag aagtctatct tttaccaatc
1981 ataatcagtg atagtggaaa tcctccactg agcagcacta gcaccttgac aatcagggtc
2041 tgtggctgca gcaatgacgg tgtcgtccag tcttgcaatg tcgaagctta tgtccttcca
2101 attggactca gtatgggcgc cttaattgcc atattagcat gcatcatttt gctgttagtc
```

2161 atcgtggtgc tgtttgtaac tctacggcgg cataaaaatg aaccattaat tatcaaagat
2221 gatgaagacg ttcgagaaaa catcattcgc tacgatgatg aaggaggagg ggaggaggac
2281 acagaggctt ttgacattgc aactttacaa aatccagatg gaattaatgg atttttaccc
2341 cgtaaggata ttaaaccaga tttgcagttt atgccaaggc aagggcttgc tccagttcca
2401 aatggtgttg atgtcgatga atttataaat gtaaggctgc atgaggcaga taatgatccc
2461 acggccccgc catatgactc cattcagata tatggctatg aaggccgagg gtcagtggct
2521 ggctccctca gctccttgga gtccaccaca tcagactcag accagaattt tgactacctc
2581 agtgactggg gtccccgctt taagagactg ggcgaactct actctgttgg tgaaagtgac
2641 aaagaaactt gacagtggat tataaataaa tcactggaac tgagcattct gtaatattct
2701 agggtcactc cccttagata caaccaatgt ggctatttgt tttagaggca agtttagcac
2761 cagtcatcta taaactcaac cacattttaa tgttgaacca aaaaaagata ataaaataaa
2821 aaagtatatg ttaggaggtt ataaatcttg tggagtgtga attaagtatg tggagtgtct
2881 agaagtcctt ggatatttga tatttacctg accaccacag acaaagatt

Figure 4

```
   1 mperlaemll dlwtpliilw itlppciyma pmnqsqvlms gsplelnslg eeqrilnrsk
  61 rgwvwnqmfv leefsgpepi lvgrlhtdld pgskkikyil sgdgagtifq indvtgdiha
 121 ikrldreeka eytltaqavd wetskplepp sefiikvqdi ndnapeflng pyhatvpems
 181 ilgtsvtnvt atdaddpvyg nsaklvysil egqpyfsiep etaiiktalp nmdreakeey
 241 lvviqakdmg ghsgglsgtt tltvtltdvn dnppkfaqsl yhfsvpedvv lgtaigrvka
 301 ndqdigenaq ssydiidgdg talfeitsda qaqdgiirlr kpldfetkks ytlkveaanv
 361 hidprfsgrg pfkdtatvki vvedadeppv fssptyllev henaalnsvi gqvtardpdi
 421 tsspirfsid rhtdlerqfn inaddgkitl atpldrelsv whnitiiate irnhsqisrv
 481 pvaikvldvn dnapefasey eaflcengkp gqiqtvsam dkddpknghy flysllpemv
 541 nnpnftikkn ednslsilak hngfnrqkqe vyllpiiisd sgnpplssts tltirvcgcs
 601 ndgvvqscnv eayvlpigls mgaliailac iilllvivvl fvtlrrhkne pliikddedv
 661 reniirydde gggeedteaf diatlqnpdg ingflprkdi kpdlqfmprq glapvpngvd
 721 vdefinvrlh eadndptapp ydsiqiygye grgsvagsls slesttsdsd qnfdylsdwg
 781 prfkrlgely svgesdket
```

Figure 5

a)

MPERLAEMLLDLWTPLIILWITLPPCIYMAPMNQSQVLMSGSPLELNSLGEEQRILNRSKRGWVWNQMFVL
EEFSGPEPILVGRVLKSVSKLH*

b)

```
      G  R  G  G  A  A  E  A  P  R  A  G  G  G  R  L  L  R  G  Q
3     ggccgcggcggtgcagcagaggcgcctcgggcaggaggagggcggcttctgcgagggcag  62
      P  E  L  H  T  D  L  D  P  G  S  K  K  I  K  Y  I  L  S  G
63    cctgagctacacacagacctggatcctgggagcaaaaaaatcaagtatatcctatcaggt  122
      D  G  A  G  T  I  F  Q  I  N  D  V  T  G  D  I  H  A  I  K
123   gatggagctgggaccatatttcaaataaatgatgtaactggagatatccatgctataaaa  182
       R  L  D  R  E  E  K  A  E  Y  T  L  T  A  Q  A  V  D  W  E
183   agacttgaccgggaggaaaaggctgagtataccctaacagctcaagcagtggactgggag  242
      T  S  K  P  L  E  P  P  S  E  F  I  I  K  V  Q  D  I  N  D
243   acaagcaaacctctggagcctccttctgaatttattattaaagttcaagacatcaatgac  302
      N  A  P  E  F  L  N  G  P  Y  H  A  T  V  P  E  ▒  S  I  L
303   aatgcaccagagtttcttaatggaccctatcatgctactgtgccagaaatgtccattttg  362
      G  T  S  V  T  N  V  T  A  T  D  A  D  D  P  V  Y  G  N  S
363   ggtacatctgtcactaacgtcactgcgaccgacgctgatgacccagtttatggaaacagt  422
      A  K  L  V  Y  S  I  L  E  G  Q  P  Y  F  S  I  E  P  E  T
423   gcaaagttggtttatagtatattggaagggcagccttatttttccattgagcctgaaaca  482
      A  I  I  K  T  A  L  P  N  ▒  D  R  E  A  K  E  E  Y  L  V
483   gctattataaaaactgcccttcccaacatggacagagaagccaaggaggagtacctggtt  542
      V  I  Q  A  K  D  ▒  G  G  H  S  G  G  L  S  G  T  T  T  L
543   gttatccaagccaaagatatgggtggacactctggtggcctgtctgggaccacgacactt  602
      T  V  T  L  T  D  V  N  D  N  P  P  K  F  A  Q  S  L  Y  H
603   acagtgactcttactgatgttaatgacaatcctccaaaatttgcacagagcctgtatcac  662
      F  S  V  P  E  D  V  V  L  G  T  A  I  G  R  V  K  A  N  D
663   ttctcagtaccggaagatgtggttcttggcactgcaataggaagggtgaaggccaatgat  722
      Q  D  I  G  E  N  A  Q  S  S  Y  D  I  I  D  G  D  G  T  A
723   caggatattggtgaaaatgcacagtcatcatatgatatcatcgatggagatggaacagca  782
      L  F  E  I  T  S  D  A  Q  A  Q  D  G  I  I  R  L  R  K  P
783   ctttttgaaatcacttctgatgcccaggcccaggatggcattataaggctaagaaaacct  842
      L  D  F  E  T  K  K  S  Y  T  L  K  V  E  A  A  N  V  H  I
843   ctggactttgagaccaaaaaatcctatacgctaaaggtagaggcagccaatgtccatatt  902
      D  P  R  F  S  G  R  G  P  F  K  D  T  A  T  V  K  I  V  V
903   gacccacgcttcagtggcaggggggccctttaaagacacggcgacagtcaaaatcgtggtt  962
      E  D  A  D  E  P  P  V  F  S  S  P  T  Y  L  L  E  V  H  E
963   gaagatgctgatgagcctccggtcttctcttcaccgacttacctacttgaagttcatgaa  1022
      N  A  A  L  N  S  V  I  G  Q  V  T  A  R
1023  aatgctgctctaaactccgtgattgggcaagtgactgctcgt etc.....
```